Vox

Why DNA test results are actually pretty fuzzy

A geneticist explains why DNA testing kits can't reveal anyone's complete ancestral history.

By Daniel Villarreal | Oct 19, 2018, 10:20am EDT

Getty Images/Westend61

You may say the 2020 presidential election cycle began this week. Sen. Elizabeth Warren (D-MA) **released the results of a DNA** test as an attempt to substantiate her claim of Cherokee ancestry — and, perhaps, to get President Donald Trump to stop mockingly calling her **"Pocahontas."**

Trump couldn't care less about the test's results, and the Cherokee Nation has since criticized the senator for publicizing them, **stating that Cherokee heritage** is determined not by DNA but by having a documented tribal ancestor.

Consumer genomic products, like those offered by 23andMe and **Ancestry.com**, have compelled more people to submit DNA samples for scientific insights into their genetic heritage, but these tests can only tell us so much. In fact, their technical limitations can affect our understanding of our genetic makeup and ancestral history.

To explore these issues, we spoke with Dr. Robert C. Green, a professor of medicine in the division of genetics at Harvard Medical School, to understand more. Green is the director of the Brigham and Women's Hospital and Broad Institute's **Genomes to People** (G2P), a research program that examines the medical, behavioral, and economic impact of using genomic information within medicine and society at large. He's also a board member of the Council for Responsible Genetics and co-founder of **Genome Medical Inc.**, a company that provides genetics expertise to people and businesses looking to understand their DNA.

Our interview has been edited for length and clarity.

Daniel Villarreal

How does genetic testing work? After someone spits into a tube and sends it off to a company like 23andMe for scientists to analyze in a lab, how do scientists process it to extract a person's genetic information?

Robert Green

It involves chemical processes to break down cells in the saliva. Sometimes spit contains cells from the inside of the cheek or cells from the glands that create spit. You just have to break open the DNA, then you centrifuge it and use other methods to isolate the DNA itself.

They then extract that DNA and analyze it by putting it into a machine. The particular DNA being examined by the chip is identified chemically, usually through a fluorescent process to mark different sites and segments. We call this chip genotyping or microarray genotyping, and it gives you, depending on the chip, somewhere in the order of a million markers all along the person's DNA.

Daniel Villarreal

How do scientists read this genetic printout to make determinations about a person's ancestry?

Robert Green

We have very accurate techniques for identifying markers, specific places along the DNA code where we know what the average person has and what other people have. For somewhere [around the] $100 price point, these chip arrays can get you a pretty accurate reading on a million or so of these markers.

What companies like **Ancestry.com**, 23andMe do is they analyze these markers in your genome and give you a reasonable sense of the probability that you have certain patterns in your DNA that indicate certain heritable traits.

When these consumer companies want to give you insights into your ancestry, they look at patterns of markers across the genome for some insights. They can look at some specific patterns on the X chromosome — remember that women have two X chromosomes, whereas men only have one — for some maternal lineage, and some of the markers at the Y chromosome, which typically men have, for some paternal lineage.

With these million markers in large numbers of people, you can do lots of different things. One of the things you can do is look at markers in different populations around the world to see different patterns. So at marker number 72 there's an A, at marker number 3,026 there's

a T, at marker number 847,000 there's a G — and that pattern is really a distinctive pattern most often seen in Asian ancestry.

You can use this in large data sets to mathematically talk about the probability that someone belongs to a particular ancestry group or is a combination of different ancestry groups, called haplogroups.

On a population basis, you can see very differences between 1,000 European Americans and 1,000 Europeans and 1,000 Africans. And you can see some differences between Americans of European ancestry and Americans of African ancestry.

The conclusions aren't perfect, and the more generations that have gone by, the more it all gets diluted and mixed, and harder and harder to interpret. Most of us ultimately are a combination of different ancestry groups. Very few of us are one pure ancestry group anymore because there's so much intermixing around the world.

Daniel Villarreal

So, roughly speaking, how did Carlos Bustamante find five segments of Native American ancestry "with very high confidence" in Elizabeth Warren's DNA? What helped him come to that conclusion?

Robert Green

Carlos Bustamante — I know him well, he's a colleague — he's a population geneticist. They study large groups of people and analyze their genetic patterns with respect to ancestry: sometimes current ethnicities, sometimes historical, sometimes archaic humans and ancient DNA and deducing how people have moved around the world and when they've interbred and when they haven't, all these things.

In very general terms, I presume he's talking about segments that are fairly distinctive to Native American populations and therefore have a highly statistical unlikeliness to be in a person's DNA unless you're related to those Native American populations.

This is a branch of science that has a lot of layers to it. Is the chip perfectly accurate? Not perfectly, no, but the chips are pretty good. Then is the interpretation accurate? There are many ways to do these interpretations. And none of them are perfectly accurate, but some of them are more well-established than others. So it's a complicated area.

I don't know how Carlos analyzed her genome. I know he's a reputable population geneticist, so if he says that there was Native American ancestry detected, I trust that he did it correctly. I'm sure he knew what kind of firestorm would follow here. But I think the whole discussion is silly, frankly.

Daniel Villarreal

Why do you think it's silly?

Robert Green

I'm glad there's some injection of science somewhere into these conversations because science is so often left outside of questions in politics. But I think the injection here into a certain sort of personal name-calling and claims and counter-claims is silly because it sidesteps the important substantive policies differences that leaders in our political system may have that influence the everyday lives of our citizens.

We should all be talking about climate change, we should be talking about America's role in the world and economic issues, and we should definitely be talking about how to improve health care. And to have the news cycles taken over by this back-and-forth around ancestry testing just seems silly.

Daniel Villarreal

What can mass consumer genomic tests like 23andMe and Ancestry.com really tell us about who we are?

Robert Green

This is a fairly multidimensional process. All of the arrays have capabilities to look at some aspects of all of these things.

On the ancestry side, they can definitely tell you who some of your relatives are. The closer the relatives are, the more accurate and specific things they can tell you. For instance, they can tell you that someone is highly likely to be your sibling, child, parent, or second cousin, or so forth.

They can tell you certain markers known to be associated with disease such as the BRCA-1 mutations that are associated with breast and ovarian cancer. They can tell you certain markers that are associated with better or worse tolerances for certain pharmaceutical medications.

They can tell you if you're carrying recessive carrier traits that put your children at some risk of birth defects and childhood diseases, and they can tell you what's called "polygenic risk scores," which is when you add up all the genes that might influence something like heart disease or Type II diabetes. You can put some people on a statistical spectrum of risk from fairly high risk to fairly low risk to mixed risk.

If your ancestry is distinctively Southeast Asian, a test can probably spot that and tell you that you're not of European background. If you had mixed ancestry for scores of generations, that's going to be much harder. So it depends on how diluted you are by ancestry mixtures.

I've heard people say, you know, "Oh! I always thought that I had Italian ancestry, but the test didn't say that." Well, maybe all the test said was that you have European background, and the Italian was already so mixed in with all the other European stuff that it didn't stand out as specifically Italian. But that doesn't mean that you're not Italian, It just means that you didn't read the part in the results that many Italian countries will be called generally European because of the mixture over multiple generations.

There are many ways to calculate your ancestry and there are many ways to interpret the information, and people should go with reputable companies and try to understand what the limitations are in the technology that they're using.

A lot of the explanations are clearly there because the best of these companies have worked extremely hard to be fair, honest, and clear. But they do depend on a customer taking their time and energy to work through the information that they have presented. I co-founded a company, Genome Medical Inc., that can help people with interpretation if they feel it's too complicated.

Daniel Villarreal

Are there concerns [about] people overrelying on these tests to tell them who they are or what their health outcomes are going to be?

Robert Green

Absolutely. And there's a vast scientific literature on this as well as a vast popular literature on this. Hundreds of thousands of newspapers articles and magazines have been written over the last 11 years since the first direct-consumer companies launched in 2007. So there's been extensive concern about misunderstanding, about overinterpretation, about false reassurance.

Genomes 2 People has written 22 different papers on how people receive their results, how they communicate their results, how they estimate or mis-estimate their results. These are all potential issues, about people misunderstanding and acting inappropriately on behalf of their genetic results.

The biggest concern on the health side is that people will somehow be falsely reassured, that if their test says, "You don't have these three mutations for breast cancer," someone will say, "Oh, good, I'm clear, I don't have any genetic concerns for breast cancer." Of course, they could have hundreds or even thousands of other migrations that aren't tested for in that particular test. Now, the company has been very clear about that — that they're just giving you these three bits, that there are plenty of others — but you have to actually read that and comprehend it to avoid that situation of false reassurance.

However, most of the research that I've performed over the last 15 years has suggested that that happens far less than people were concerned about. People generally do understand very well, they generally do not get terribly distressed, and they generally do not do medically inappropriate things. I'm sure we can find exceptions, but our research has suggested that these are not large-scale dangers.

Daniel Villarreal

How have these home consumer genomic tests changed the ways people interpret their family history and their identity?

Robert Green

I think it's in the process of really opening people's eyes about the nature of human genetics, how similar we are, how much our beings have mixed over the millennia.

I think it's teaching people that most genetics is not deterministic — that these markers don't mean you're definitely all one ancestry or you're definitely gonna have a disease — but that most genetics is probabilistic, meaning that we measure the *likelihood* that something is true.

And I think it's teaching people about the complexity of genetics, which is really important because some companies — not the leading ones — but some companies really try to oversimplify this.

I think we need a world where the citizenry has enough understanding of some basic genetic principles that they can be intelligent consumers. One of the things I like to say, especially on the consumer side of medical genetics, is that people have a right to this information. I believe — some of my colleagues feel differently — these companies have a right to promote and sell it.

But I do believe that people take on an additional responsibility to read what the company tells them on the website — to think about it, not simply to glance at it and invite misunderstanding — because it is an interesting world. It's one that people are deeply interested in, but whether you're talking about ancestry or medical genetics, it's not intuitively straightforward all the time.